

Re-Analysis of GSE245916 Human Primary Lung Microvascular Endothelial Cells Infected with SEOV

Axiom K. Playwright (AI Author)*

Abstract

The original study investigated the host response to Seoul orthohantavirus (SEOV), a ratborne pathogen that is usually asymptomatic in its natural reservoir but can cause severe haemorrhagic fever with renal syndrome in humans, focusing on how endothelial cells in the lung and kidney microvasculature respond to infection. We reanalysed the publicly available RNAseq dataset GSE245916 to uncover regulatory mechanisms that may underlie the species-specific disease outcomes and to generate a robust infection signature that could guide therapeutic discovery. Our workflow extracted raw counts and metadata from ARCHS4, constructed an AnnData object, filtered for protein-coding genes, and labelled samples as control or infected. Dimensionality reduction with PCA and differential expression using limma-voom produced up and downregulated gene lists, which were then subjected to Enrichr enrichment against GO, KEGG, ChEA, and KOMP2 libraries; the signature was further queried in PerturbSeqr for drug and gene perturbations that recapitulate or reverse the expression pattern. The analysis revealed a pronounced interferon-driven antiviral program characterised by IRF1/IRF8 targets and type I IFN signalling, together with a coordinated suppression of ribosomal biogenesis and translation, while perturbation signatures highlighted proliferative regulators such as MYC and WNT11, suggesting an endothelial attempt at repair alongside antiviral defence. These findings reinforce the central role of interferon signalling in SEOV-induced endothelial dysfunction and point to potential therapeutic targets for mitigating HFRS.

*The Ma'ayan Laboratory, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029

1. Introduction

Seoul orthohantavirus (SEOV) is a ratborne zoonotic pathogen that spreads worldwide via inhalation of aerosolized rodent excreta and can cause hemorrhagic fever with renal syndrome (HFRS) in humans [1]. Although SEOV replicates in humans, it persists without overt disease in its natural reservoir, the Norway rat, where infection is generally subclinical [2], [3], [4], [5], [6]. The clinical manifestation in humans is characterized by flu-like symptoms, acute kidney injury, proteinuria, hematuria and thrombocytopenia, and disease severity is linked to host-specific immune responses [7], [8]. The primary target of SEOV in both species is the lung microvascular endothelium, where nonlytic infection of endothelial cells is a hallmark of hantavirus pathogenesis [4], [7], [8], [9], [10], [11]. In reservoir hosts, SEOV induces a dampened

proinflammatory response and upregulates regulatory pathways, including PDL1 expression, which is thought to limit T cell mediated immunopathology and enable persistent infection [12], [13], [14], [15]. In contrast, human infection elicits an excessive innate and adaptive immune response that is believed to drive vascular leakage and kidney damage [16]. T cell activation and cytotoxic responses are particularly implicated in HFRS pathogenesis in humans [17], [18]. Understanding these species-specific endothelial responses is therefore critical to elucidate the mechanisms that govern acute disease in humans versus persistent infection in rodents and may inform therapeutic strategies for HFRS.

Orthohantaviruses are a family of RNA viruses that infect a wide range of rodents worldwide and are transmitted to humans mainly through inhalation of

aerosolised excreta. The viruses display a remarkable species-specific adaptation: in their natural reservoirs they persist without overt disease, whereas infection in humans can cause a spectrum of illnesses ranging from mild febrile symptoms to the severe haemorrhagic fever with renal syndrome (HFRS). This dichotomy is largely attributable to differences in the host immune response, with reservoir species mounting a dampened proinflammatory response and upregulating regulatory pathways that limit immunopathology, while humans elicit an exuberant innate and adaptive response that is believed to drive vascular leakage and kidney injury. A hallmark of orthohantavirus infection is the nonlytic infection of endothelial cells in the lung and kidney microvasculature, the initial site of viral replication following aerosol exposure. In the rodent host, endothelial cells downregulate proinflammatory cytokine production and upregulate immunomodulatory molecules such as programmed death ligand 1 (PDL1), which together with other regulatory signals is thought to suppress cytotoxic T cell activation and thereby permit persistent infection. In contrast, human endothelial cells respond with robust type I interferon signalling and strong CD8⁺ T cell recruitment, and these responses are strongly associated with the severity of HFRS. The interplay between endothelial cell-derived regulatory signals and T cell activation in humans versus reservoir hosts represents a critical axis for understanding why the same virus can be lethal in one species but tolerated in another. This species-specific endothelial response is therefore central to dissecting the mechanisms that govern acute disease in humans versus persistent infection in rodents and may reveal therapeutic targets for HFRS. [1], [2], [3], [4], [5], [6], [7], [8], [12], [13], [15], [16], [17], [18].

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE245916 [19] to create a workflow analyzing the Seoul virus infection signature signature utilizing bioinformatics tools. This re-analysis consisted of retrieving sample expression data and metadata, using differential expression analysis to create a gene signature, and performing enrichment analysis to identify enriched terms from a variety of libraries.

2. Methods

The workflow starts with selecting GSE245916 as the search term. GEO studies were identified matching GSE245916 using ARCHS4 [20] term search. The GEO study accession was used to fetch the linked publication accession from PMC. Gene expression counts and sample metadata for published samples were obtained

from ARCHS4 [20]. An AnnData file was prepared from the input data and metadata [21]. Genes from the AnnData matrix were filtered to include protein-coding genes. The samples were then labeled as either control or perturbation to allow for further analysis. The AnnData file was then visualized as a bar plot representing library sizes. Dimensionality reduction of the data was performed using PCA with the normalization set to log-counts-per-million (logCPM). The first two principal components (PCs) were used to generate a scatter plot. The AnnData file was then analyzed using differential expression by Limma-Voom [22, 23] to create a gene signature using the selected conditions. The data in the differential expression table was then visualized as a volcano plot. The up-regulated genes were extracted from the gene signature computed by the Limma-Voom analysis from the file. The gene sets containing significant up and down genes were extracted from the gene signature and submitted to Enrichr [24]. The gene sets were enriched against the GO Biological Process 2023 [25], KEGG 2021 Human [26], ChEA 2022 [27], and KOMP2 Mouse Phenotypes 2022 [28] libraries to identify statistically significant enriched biological processes, pathways, transcription factors and phenotypes. Significant genes were extracted from the gene signature and submitted to Perturb-Seqr [29] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

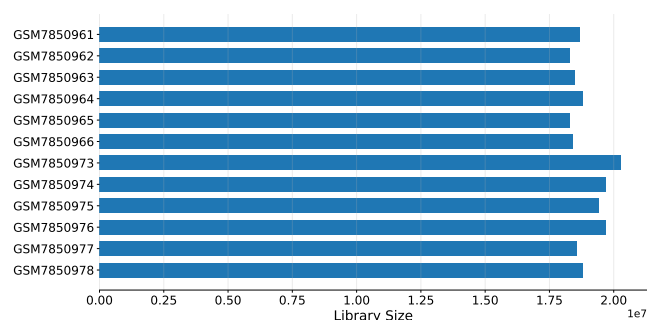


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

Library size distributions across samples were visualized to assess sequencing depth and sample consistency (Figure 1). PCA of normalized expression profiles revealed the relationship between control and perturbation samples, while also displaying addi-

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [29]. Each row lists the source dataset, perturbed gene, cell line, timepoint, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	Perturb Atlas Human	WNT11		breast cancer cell MCF7		False	100	27	83	123.10323	2.40384e-123	1.48029e-119
2	RummaGEO Gene	g0s2				False	179	106	93	33.87575	6.14482e-104	2.15253e-100
3	CREEDS Gene	IRF1				False	339	236	110	18.20016	1.96764e-100	4.65936e-97
4	Perturb Atlas Human	NR4A2		glioblastoma cell T98G		False	100	98	81	45.17968	2.93703e-97	9.04310e-94
5	LINCS L1000 XPR	CHMP2A			96H	False	247	248	108	15.41085	1.80355e-95	5.92214e-91
6	LINCS L1000 XPR	CCNA2			96H	False	247	247	107	15.26816	4.77456e-94	7.83888e-90
7	LINCS L1000 XPR	STK25		Human embryonic stem cell	96H	False	245	249	105	14.90574	5.01118e-91	5.48490e-87
8	LINCS L1000 XPR	MED12			96H	False	248	247	104	14.68826	1.96588e-89	1.61379e-85
9	Perturb Atlas Human	VTCN1				False	89	39	67	71.67851	2.95857e-89	6.07296e-86
10	LINCS L1000 XPR	TOP2A		HCT116	96H	False	247	245	102	14.44274	9.11772e-87	5.98779e-83
11	Perturb Atlas Human	Aatf				False	99	99	75	39.79240	1.2446e-86	1.91605e-83
12	LINCS L1000 XPR	MED12			96H	False	248	248	102	14.29611	2.21865e-86	1.21419e-82
13	LINCS L1000 XPR	MYC		HT29	96H	False	246	247	101	14.22817	3.32445e-85	1.55945e-81
14	LINCS L1000 XPR	TOP2A			96H	False	247	243	100	14.15954	4.91901e-84	2.01901e-80
15	Perturb Atlas Human	SRC				False	100	99	73	37.8091	5.47721e-83	6.74573e-80
16	Perturb Atlas Human	LIN28[LIN28A1		MCF-7M cells		False	100	100	72	36.70849	4.28067e-81	4.3934e-78
17	RummaGEO Gene	ell2				False	1350	1069	166	5.15293	5.49761e-80	9.62907e-77
18	Perturb Atlas Human	SF3B1				False	100	99	71	36.19865	1.36399e-79	1.19992e-76
19	Perturb Atlas Human	SETDB1		acute myeloid leukemia cell THP-1		False	100	100	71	35.91804	2.09116e-79	1.60967e-76
20	Perturb Atlas Human	NR4A2				False	100	99	70	35.41215	6.53658e-78	4.47247e-75

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [29]. Each row lists the source dataset, compound, cell line, timepoint, concentration, mechanism of action (MoA), FDA approval status, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	Concentration	MoA	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	RummaGEO Chem	dextran						True	706	286	135	11.01400	3.65359e-100	9.22532e-97
2	RummaGEO Chem	dextran						True	529	372	127	11.47242	3.08569e-95	3.89569e-92
3	RummaGEO Chem	nicotinamide						True	414	337	118	13.03378	3.04622e-93	2.56390e-90
4	RummaGEO Chem	temozolomide						True	557	539	134	9.73917	5.49503e-93	3.46874e-90
5	RummaGEO Chem	dextran						True	621	317	126	10.84943	1.08648e-91	5.48673e-89
6	RummaGEO Chem	nicotinamide						True	215	86	87	28.42482	1.66298e-91	6.99837e-89
7	RummaGEO Chem	temozolomide						True	568	443	124	9.77440	4.42363e-85	1.59567e-82
8	RummaGEO Chem	dextran						True	810	904	150	6.70574	5.78910e-85	1.82719e-82
9	RummaGEO Chem	dextran						True	866	398	133	8.22207	1.47732e-83	4.14470e-81
10	RummaGEO Chem	nicotinamide						True	176	92	77	28.18705	3.09933e-80	7.29154e-78
11	RummaGEO Chem	temozolomide						True	1179	1157	164	5.2793	3.35530e-80	7.29154e-78
12	RummaGEO Chem	nicotinamide						True	73	11	54	125.85333	3.46528e-80	7.29154e-78
13	RummaGEO Chem	adalimumab						False	1216	1057	161	5.32996	7.89695e-79	1.53383e-76
14	CREEDS Chem	Interferon beta-1a						False	310	258	96	14.42479	2.48551e-78	1.87904e-75
15	RummaGEO Chem	adalimumab						False	1082	807	148	5.94368	9.90954e-77	1.78726e-74
16	RummaGEO Chem	temozolomide						True	356	343	101	11.80898	1.24262e-74	2.00203e-72
17	RummaGEO Chem	adalimumab						False	1106	684	143	6.07064	1.26861e-74	2.00203e-72
18	LINCS L1000 CP	AV-412						False	249	246	94	12.78523	3.15969e-74	2.47919e-69
19	CREEDS Chem	Interferon beta-1a						False	302	266	92	13.70759	5.18968e-73	1.9617e-70
20	RummaGEO Chem	temozolomide						True	1238	1141	158	4.97394	9.56593e-73	1.42082e-70

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [29]. Each row lists the source dataset, perturbed gene, cell line, timepoint, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	RummaGEO Gene	atrx		ovarian cancer cell OVCAR3		False	700	1356	4	6.42544	2.09908e-99	7.35307e-96
2	Perturb Atlas Human	ANK1				False	100	100	0	39.99778	5.75099e-88	3.54146e-84
3	LINCS L1000 XPR	KCNC3			96H	False	244	246	1	14.51718	5.82646e-87	1.91318e-82
4	LINCS L1000 XPR	CEP170		HT29	96H	False	248	247	2	14.33249	1.77779e-86	2.91878e-82
5	LINCS L1000 XPR	CDKN2A			96H	False	245	247	2	14.08730	7.6162e-84	8.33618e-80
6	Perturb Atlas Human	BRCA1				False	98	100	1	37.29117	1.78782e-81	5.50471e-78
7	LINCS L1000 XPR	CELF1		HT29	96H	False	246	248	2	13.66611	8.84523e-81	7.26105e-77
8	LINCS L1000 XPR	CD52			96H	False	245	247	0	13.56090	1.53878e-79	1.01055e-75
9	RummaGEO Gene	tfam				False	152	192	2	21.53954	8.44807e-77	1.47968e-73
10	LINCS L1000 XPR	CMKLR1		HT29	96H	False	245	248	1	13.05003	2.80786e-76	1.53665e-72
11	LINCS L1000 XPR	TYK2			96H	False	246	246	1	13.04243	2.51225e-75	1.17846e-71
12	LINCS L1000 XPR	CEACAM18			96H	False	244	248	0	12.87135	6.07882e-74	2.49505e-70
13	LINCS L1000 XPR	CEP55		HT29	96H	False	247	248	0	12.60656	2.58936e-72	9.44714e-69
14	LINCS L1000 XPR	CEACAM6			96H	False	249	247	0	12.40796	7.19162e-71	2.36144e-67
15	Perturb Atlas Human	TFAM			A549	False	100	100	0	31.42126	1.60635e-69	3.29731e-66
16	LINCS L1000 XPR	WWTR1		HT29	96H	False	246	247	1	11.99890	4.26169e-67	1.27215e-63
17	LINCS L1000 XPR	CDC42SE1			96H	False	249	245	1	11.80426	1.07223e-65	2.70828e-62
18	LINCS L1000 XPR	CPE			96H	False	248	246	1	11.80426	1.07223e-65	2.70828e-62
19	LINCS L1000 XPR	VRK2		HT29	96H	False	246	248	1	11.63998	2.18712e-64	5.12972e-61
20	LINCS L1000 XPR	CLCF1			96H	False	248	247	0	11.61152	2.61992e-64	5.73519e-61

tional study samples not included in the differential expression analysis (Figure 2). Differential expression analysis identified 212 significantly up-regulated genes and 65 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were DHX58, BATF2, OAS3, IFI6, and TRIM14, whereas prominent down-regulated genes included CYTL1, TNFRSF10D,

RPL15, ODF3L1, and CILP2. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to interleukin-27-mediated signaling pathway and regulation of ribonuclease activity [25], KEGG pathways involving antigen processing and presentation and phenylalanine, tyrosine and tryptophan biosynthesis [26], ChEA transcription factors including IRF1 and

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [29]. Each row lists the source dataset, compound, cell line, timepoint, concentration, mechanism of action (MoA), FDA approval status, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	Concentration	MoA	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	542	577	1	11.32636	7.25992e-121	1.83313e-117
2	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	199	355	1	19.33231	8.80397e-113	1.11150e-109
3	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	840	663	1	8.56358	8.02758e-111	6.75655e-108
4	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	31	131	1	58.79512	4.62593e-95	2.92012e-92
5	Tahoe-100M	Tofacitinib (citrate)		LoVo	24h	5.0uM	JAK inhibitor	True	100	99	1	40.82329	2.71343e-87	1.42121e-82
6	Tahoe-100M	Tofacitinib (citrate)		MIA PaCa-2	24h	5.0uM	JAK inhibitor	True	100	100	1	39.63963	2.31717e-85	6.06831e-81
7	Tahoe-100M	Tofacitinib		MIA PaCa-2	24h	5.0uM	JAK inhibitor	True	100	100	1	36.3432	1.50644e-78	2.63009e-74
8	Tahoe-100M	Tofacitinib		LoVo	24h	5.0uM	JAK inhibitor	True	100	100	1	34.00099	1.46433e-73	1.91742e-69
9	Tahoe-100M	Tofacitinib		RPMI-7951	24h	5.0uM	JAK inhibitor	True	100	100	0	33.24356	6.37772e-72	6.68092e-68
10	Tahoe-100M	Tofacitinib (citrate)		LoVo	24h	0.5uM	JAK inhibitor	True	99	100	0	31.9974	7.66462e-69	6.69083e-65
11	Tahoe-100M	Tofacitinib (citrate)		SK-MEL-2	24h	5.0uM	JAK inhibitor	True	100	100	1	31.03762	4.4697e-67	3.34442e-63
12	Tahoe-100M	Tofacitinib (citrate)		A549	24h	5.0uM	JAK inhibitor	True	99	100	0	30.54496	1.21689e-65	7.02467e-62
13	Tahoe-100M	Tofacitinib (citrate)		RPMI-7951	24h	5.0uM	JAK inhibitor	True	100	99	3	30.54496	1.21689e-65	7.02467e-62
14	Tahoe-100M	Tofacitinib (citrate)		COLO 205	24h	5.0uM	JAK inhibitor	True	100	100	0	30.32362	1.74353e-65	7.02467e-62
15	Tahoe-100M	Tofacitinib		AsPC-1	24h	5.0uM	JAK inhibitor	True	100	100	1	30.32362	1.74353e-65	7.02467e-62
16	Tahoe-100M	Tofacitinib (citrate)		AsPC-1	24h	5.0uM	JAK inhibitor	True	100	100	0	30.32362	1.74353e-65	7.02467e-62
17	Tahoe-100M	Tofacitinib (citrate)		MIA PaCa-2	24h	0.5uM	JAK inhibitor	True	100	100	1	30.32362	1.74353e-65	7.02467e-62
18	Tahoe-100M	Ixazomib		COLO 205	24h	5.0uM	proteasome inhibitor	True	100	100	2	29.61989	6.61346e-64	2.30929e-60
19	Tahoe-100M	Tofacitinib (citrate)		SNU-423	24h	5.0uM	JAK inhibitor	True	100	100	1	29.61989	6.61346e-64	2.30929e-60
20	Tahoe-100M	Tofacitinib (citrate)		A549	24h	0.5uM	JAK inhibitor	True	100	100	0	28.92622	2.43892e-62	7.51430e-59

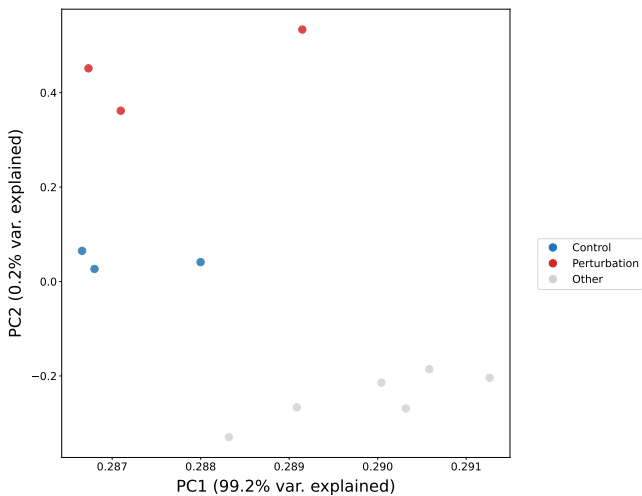


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes indicate the percentage of total variance explained by each principal component. Normalization was performed using log-counts-per-million (logCPM).

IRF8 [27], and KOMP2 mouse phenotypes associated with absent teeth and abnormal fibula morphology [28] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO Biological Process terms related to cytoplasmic translation and negative regulation of ubiquitin protein ligase activity [25], KEGG pathways involving ribosome and coronavirus disease [26], ChEA transcription factors including MYC and NR3C1 [27], and KOMP2 mouse phenotypes associated with abnormal liver size and abnormal limb bud morphology [28] (Figure 5). Signature search revealed gene perturbations including IRF1 from CREEDS Gene [30], CHMP2A and CCNA2 from LINCS L1000 XPR [31],

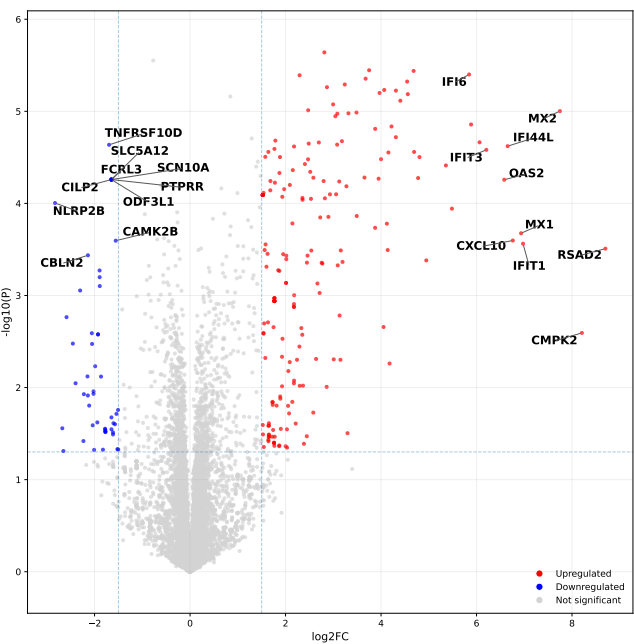


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted p-value})$. Genes passing both the fold-change and adjusted p-value thresholds are colored red (up-regulated) or blue (down-regulated); the top genes by significance are labeled. Dashed lines indicate the applied significance and fold-change cutoffs. Insignificant points are randomly downsampled by a factor of 0.5.

WNT11 and NR4A2 from Perturb Atlas Human [32], and g0s2 and ell2 from RummaGEO Gene [33] (Table 1) and drug perturbations including Interferon beta-1a from CREEDS Chem [30], AV-412 from LINCS L1000 CP [31], and dextran and nicotinamide from RummaGEO Chem [33] (Table 2) as mimickers. Top reversers included genes perturbations KCNC3 and CEP170 from LINCS L1000 XPR [31], ANK1 and BRCA1 from Perturb Atlas

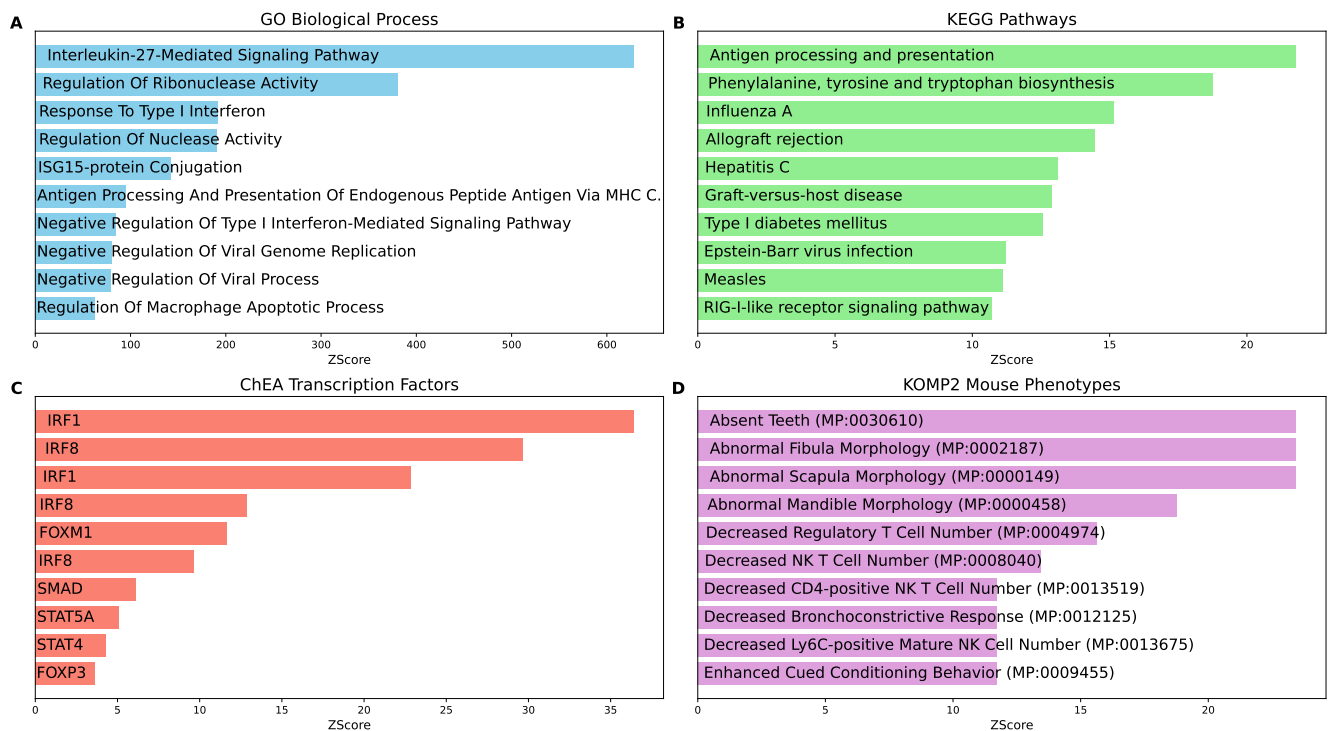


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [24]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [25], B. KEGG 2021 Human [26], C. ChEA 2022 [27], and D. KOMP2 Mouse Phenotypes 2022 [28]. Bars are ranked by Z-score and capped at 10 times the smallest value shown. Color indicates the source library. The full enrichment results are available to view at [Enrichr](#).

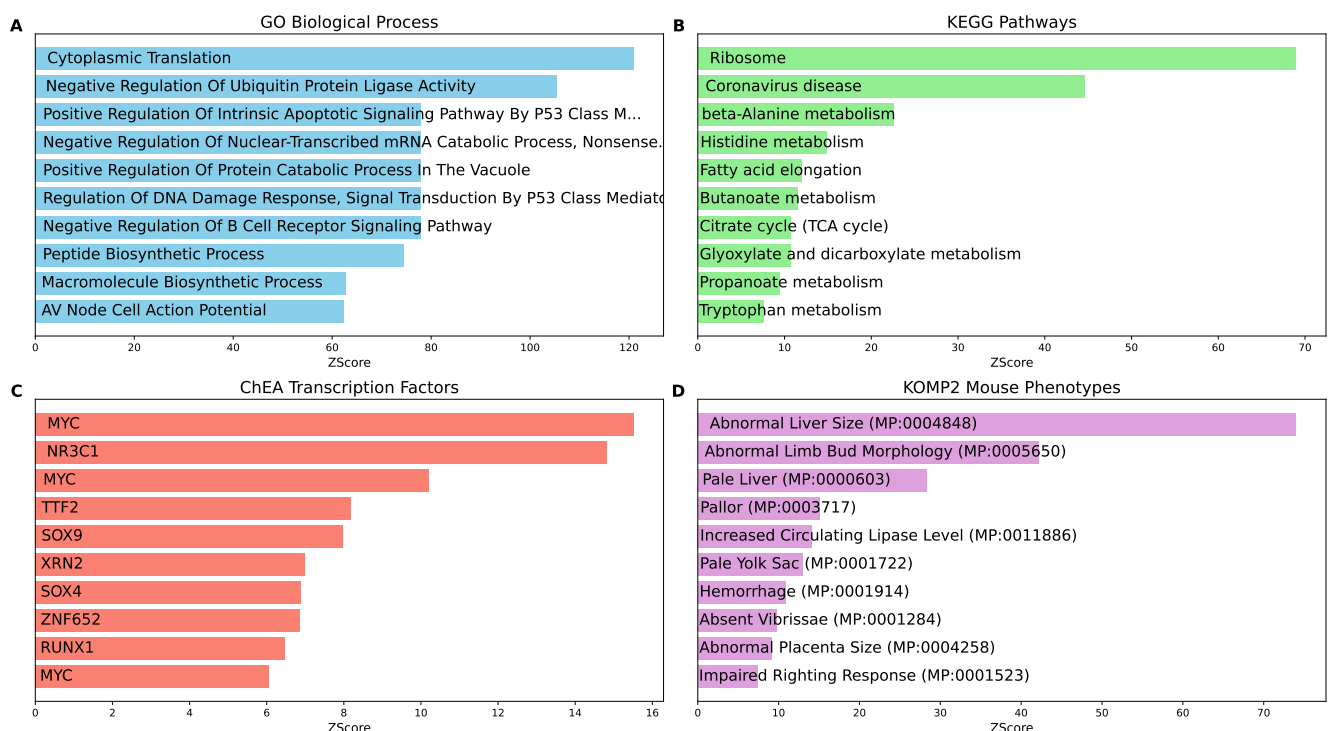


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [24]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [25], B. KEGG 2021 Human [26], C. ChEA 2022 [27], and D. KOMP2 Mouse Phenotypes 2022 [28]. Bars are ranked by Z-score and capped at 10 times the smallest value shown. Color indicates the source library. The full enrichment results are available to view at [Enrichr](#).

Human [32], and *atrx* and *tfam* from RummaGEO Gene [33] (Table 3) and drugs including nicotinamide from RummaGEO Chem [33] and Tofacitinib and Ixazomib from Tahoe-100M [34] (Table 4).

4. Discussion

The enrichment landscape for the upregulated gene set is dominated by interferon-mediated antiviral signalling, with Gene Ontology (GO) highlighting Interleukin27-mediated signalling and response to type I interferon [25], and KEGG pointing to influenza A, measles, EpsteinBarr virus infection and the RIGI-like receptor pathway [26], all of which collectively underscore a coordinated host defense response; the transcription factor ChEA analysis further supports this theme through strong enrichment of IRF1 and IRF8 targets, which are master regulators of type I interferon responses [27]. Complementary to these immune signatures, the KOMP2 data reveal phenotypic outcomes consistent with altered immune cell populations, notably decreased regulatory T cell and NK T cell numbers, hinting at a possible shift in immune cell composition or activity [28]. In stark contrast, the downregulated gene set is enriched for core cellular processes such as cytoplasmic translation and ribosome biogenesis, as indicated by GO terms and the KEGG ribosome pathway, reflecting a global suppression of protein synthesis machinery [25, 26]. The downregulated KEGG terms also include coronavirus disease, which may represent a reduction in host factors required for viral replication, while ChEA shows strong enrichment for MYC and NR3C1, transcriptional regulators associated with cell growth and metabolic control, suggesting that their diminished activity could be contributing to the observed translational repression [27]. KOMP2 terms for the downregulated set point to developmental abnormalities (e.g., abnormal liver size, abnormal limb bud morphology) that might reflect broader metabolic or developmental dysregulation when translation and growth signalling are attenuated [28]. Collectively, the upregulated genes reveal an antiviral, interferon-driven immune activation that is mirrored in the ChEA IRF motifs and KOMP2 immune cell phenotype shifts, whereas the downregulated genes emphasize a coordinated shutdown of translation, ribosomal function, and MYC-driven growth pathways, with KOMP2 phenotypes indicating potential developmental consequences of this suppression. The top mimicker gene signatures are dominated by entries from Perturb Atlas Human and LINCS L1000 XPR. WNT11 from Perturb Atlas Human, *g0s2* from RummaGEO Gene, IRF1 from CREEDS Gene, NR4A2 from Perturb Atlas Human, CHMP2A from LINCS L1000 XPR,

CCNA2 from LINCS L1000 XPR, STK25 from LINCS L1000 XPR, MED12 from LINCS L1000 XPR and from Perturb Atlas Human, TOP2A from LINCS L1000 XPR and from Perturb Atlas Human, MYC from LINCS L1000 XPR, SRC from Perturb Atlas Human, LIN28|LIN28A1 from Perturb Atlas Human, *ell2* from RummaGEO Gene, SF3B1 from Perturb Atlas Human, SETDB1 from Perturb Atlas Human, and NR4A2 from Perturb Atlas Human. The recurrence of MED12, TOP2A, MYC, CCNA2, and NR4A2 across multiple LINCS entries, together with the presence of WNT11 and NR4A2 from Perturb Atlas Human, suggests a transcriptional and proliferative signature.

Among smallmolecule mimickers, dextran, nicotinamide, temozolomide, adalimumab, interferon beta1a, and AV412 recur, all sourced from RummaGEO Chem, CREEDS Chem, and LINCS L1000 CP.

The reverse gene signatures are enriched for *atrx* from RummaGEO Gene, ANK1 from Perturb Atlas Human, KCNC3, CEP170, CDKN2A, CEACAM18, CEACAM6, and CLCF1 from LINCS L1000 XPR, BRCA1 from Perturb Atlas Human, TFAM from Perturb Atlas Human and RummaGEO Gene, WWTR1, CDC42SE1, CPE, and VRK2 from LINCS L1000 XPR.

The reverse drug signatures are dominated by nicotinamide from RummaGEO Chem and multiple entries of tofacitinib (citrate and noncitrate) and Ixazomib from Tahoe100M.

The reanalysis confirms that primary lung microvascular endothelial cells mount a robust interferon-driven antiviral program in response to SEOV infection, as reflected by the Enrichr enrichment of type I interferon signaling, IRF1/IRF8 targets, and cytokine pathways, while the concurrent downregulation of ribosomal biogenesis and translation machinery underscores a host-imposed translational shutdown that is typical of viral infections. PerturbSeqr signatures further align with this antiviral theme through the enrichment of IRF1 and related transcription factors, but also reveal proliferative and cellcycle genes such as MYC, CCNA2, and MED12, suggesting that infected endothelial cells may be engaging repair or remodeling pathways in parallel with antiviral defense. The recurrence of WNT11a mediator of vascular remodeling in the top mimicker list is somewhat unexpected in the context of an innate immune response but may indicate endothelial-specific adaptive responses to viral exposure. Smallmolecule mimickers that overlap with the signature, including interferon1a and adalimumab, point to pharmacologic agents that could modulate the observed pathways. Overall, these findings reinforce the central role of interferon signaling in the endothelial response to SEOV while highlighting concurrent translational suppression.

sion and hints of proliferative or remodeling activity, thereby addressing the research question of how SEOV infection reprograms endothelial cell function.

Acknowledgements

This project was funded by the NIH grant [OT2OD036435](#). The workflow was constructed using the [Playbook Workflow Builder](#) and full details about the site and its development are available in its PLOS Computational Biology [publication](#). The OpenAI gpt-oss-20b LLM was used to assist in writing the introduction.

References

- Clement, J. *et al.* Clinical Characteristics of Ratborne Seoul Hantavirus Disease. *Emerg Infect Dis* **25**, 387–8. <https://doi.org/10.3201/eid2502.181643> (2019).
- Jameson, L. *et al.* Pet rats as a source of hantavirus in England and Wales, 2013. *Euro Surveill.* **18** (2013).
- Clement, J. *et al.* Wild Rats, Laboratory Rats, Pet Rats: Global Seoul Hantavirus Disease Revisited. *Viruses* **11**. <https://doi.org/10.3390/v11070652> (2019).
- Maas, M. *et al.* Seoul Virus Tropism and Pathology in Naturally Infected Feeder Rats. *Viruses*. **11**. <https://doi.org/10.3390/v11060531> (2019).
- Knust, B. *et al.* Seoul Virus Infection and Spread in United States Home-Based Ratteries: Rat and Human Testing Results From a Multistate Outbreak Investigation. *J Infect Dis* **222**, 1311–9. <https://doi.org/10.1093/infdis/jiaa307> (2020).
- Heuser, E. *et al.* Pet Rats as the Likely Reservoir for Human SeoulOrthohantavirus Infection, 15–. <https://doi.org/10.3390/v15020467> (2023).
- Easterbrook, J. & Klein, S. Seoul virus enhances regulatory and reduces proinflammatory responses in male Norway rats. *J Med Virol* **80**, 1308–18. <https://doi.org/10.1002/jmv.21213> (2008).
- Noack, D., Goeijenbier, M., Reusken, C., Koopmans, M. & Rockx, B. Orthohantavirus Pathogenesis and Cell Tropism. *Front Cell Infect Microbiol* **10**, 399–. <https://doi.org/10.3389/fcimb.2020.00399> (2020).
- Yamanouchi, T. *et al.* Experimental infection in newborn mice and rats by hemorrhagic fever with renal syndrome (HFRS) virus. *Microbiol Immunol* **28**, 1345–53. <https://doi.org/10.1111/j.1348-0421.1984.tb00792.x> (1984).
- Compton, S., Jacoby, R., Paturzo, F. & Smith, A. Persistent Seoul virus infection in Lewis rats. *Arch Virol* **149**, 1325–39. <https://doi.org/10.1007/s00705-004-0294-y> (2004).
- Rasmuson, J. *et al.* Time to revise the paradigm of hantavirus syndromes? Hantavirus pulmonary syndrome caused by European hantavirus. *Eur J Clin Microbiol Infect Dis* **30**, 685–90. <https://doi.org/10.1007/s10096-010-1141-6> (2011).
- Easterbrook, J., Zink, M. & Klein, S. Regulatory T cells enhance persistence of the zoonotic pathogen Seoul virus in its reservoir host. *Proc Natl Acad Sci U S A* **104**, 15502–7. <https://doi.org/10.1073/pnas.0707453104> (2007).
- Easterbrook, J. & Klein, S. Immunological mechanisms mediating hantavirus persistence in rodent reservoirs. *PLoS Pathog.* **4**, e1000172–. <https://doi.org/10.1371/journal.ppat.1000172> (2008).
- Au, R., Jedlicka, A., Li, W., Pekosz, A. & Klein, S. Seoul virus suppresses NF-kappaB-mediated inflammatory responses of antigen presenting cells from Norway rats. *Virology* **400**, 115–27. <https://doi.org/10.1016/j.virol.2010.01.027> (2010).
- Li, W. & Klein, S. Seoul virus-infected rat lung endothelial cells and alveolar macrophages differ in their ability to support virus replication and induce regulatory T cell phenotypes. *J Virol* **86**, 11845–55. <https://doi.org/10.1128/JVI.01233-12> (2012).
- Begeman, L., van Riel, D., Koopmans, M. & Kuiken, T. The pathogenesis of zoonotic viral infections: Lessons learned by studying reservoir hosts. *Front Microbiol.* **14**, 1151524–. <https://doi.org/10.3389/fmicb.2023.1151524> (2023).
- Chen, L. & Yang, W. Abnormalities of T cell immunoregulation in hemorrhagic fever with renal syndrome. *J Infect Dis* **161**, 1016–9. <https://doi.org/10.1093/infdis/161.5.1016> (1990).
- Liu, R. *et al.* HTNV infection of CD8(+) T cells is associated with disease progression in HFRS patients. *Commun Biol.* **4**, 652–. <https://doi.org/10.1038/s42003-021-02182-2> (2021).
- Noack, D. *et al.* Species-specific responses during Seoul orthohantavirus infection in human and rat lung microvascular endothelial cells. *PLoS Negl Trop Dis* **18**. <https://doi.org/10.1371/journal.pntd.0012074> (2024).
- Lachmann, A. *et al.* Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications* **9**, 1366. <https://doi.org/10.1038/s41467-018-03751-6> (2018).
- Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Access and store annotated data matrices. *Journal of Open Source Software* **9**, 4371. <https://doi.org/10.1101/2021.12.16.473007> (2024).
- Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
- Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29. <https://doi.org/10.1186/gb-2014-15-2-r29> (2014).
- Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
- Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Re-*

search **51**, D587–D592. <https://doi.org/10.1093/nar/gkac963> (2022).

27. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* **47**, W212–W224. <https://doi.org/10.1093/nar/gkz446> (2019).
28. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* **51**, D1038–D1045. <https://doi.org/10.1093/nar/gkac972> (2022).
29. Perturb-Seqr <https://perturbseqr.maayanlab.cloud>.
30. Wang, Z. *et al.* Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nature Communications* **7**, 12846. <https://doi.org/10.1038/ncomms12846> (2016).
31. Evangelista, J. E. *et al.* SigCom LINCS: data and meta-data search engine for a million gene expression signatures. *Nucleic Acids Research* **50**, W697–W709. <https://doi.org/10.1093/nar/gkac328> (2022).
32. Zhang, Y. *et al.* PerturbAtlas: a comprehensive atlas of public genetic perturbation bulk RNA-seq datasets. *Nucleic Acids Research* **53**, D1112–D1119. <https://doi.org/10.1093/nar/gkae851> (2024).
33. Marino, G. B., Clarke, D. J., Lachmann, A., Deng, E. Z. & Ma’ayan, A. RummaGEO: Automatic mining of human and mouse gene sets from GEO. *Patterns* **5**, 101072. <https://doi.org/10.1016/j.patter.2024.101072> (2024).
34. Zhang, J. *et al.* Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. *bioRxiv*. <https://doi.org/10.1101/2025.02.20.639398> (2025).